Can Local Avatars Satisfy A Global Audience? A Case Study of High-Fidelity 3D Facial Avatar Animation in Subject Identification and Emotion Perception by US and International Groups

CHANG YUN, ZHIGANG DENG, and MERRILL HISCOCK
University of Houston

This study investigates effectiveness of a local high-fidelity 3D facial avatar for a global audience by observing how US and International student groups differed in identifying subjects and perceiving emotions while viewing nonverbal high-fidelity 3D facial avatar animations embedded with the motion data of three US individuals. To synthesize the animated 3D avatars to convey highly believable facial expressions, a 3D scanned facial model was mapped with high-fidelity motion-capture data of three native US subjects as they spoke designated English sentences with specified emotions. Simple animations in conjunction with actual footage of the subjects speaking during the facial motion-capture sessions were shown several times to both native US and international students in similar settings. After a familiarization process, we showed the students randomly arranged talking avatars without voices and asked them to identify the corresponding identities and emotional types of the subjects whose facial expressions were utilized in the creation of the avatars, and to rate their confidence in their selections. We found that the US group had higher success rates in subject identification, although the related difference in confidence ratings between two groups was not significant. The differences in the success rates and confidence ratings on the perception of emotion between the two groups were not significant either. The results of our study provide interesting insights into avatar-based interaction where the national and/or cultural background of a person impacts the perception of identity while having little effect on the perception of emotion. However, we observed that dynamics (e.g., head motion) could offset the disadvantage of cultural unfamiliarity in subject identification. We observed that both groups performed at a nearly identical level in subject identification and emotion perception when they were shown the avatar animation with heightened expression and dynamic intensities. In addition, we observed that the confidence ratings were correlated to accuracy in identifying the subject but not to accuracy in perceiving emotion.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Preservation**]: User Interfaces—*Benchmarking; Evaluation / methodology; Interaction styles*

General Terms: Human Factors

Authors' address: University of Houston; contact author's email address: Chang.Yun@mail.uh.edu.

DOI 10.1145/1541895.1541901 http://doi.acm.org/10.1145/1541895.1541901

Additional Key Words and Phrases: Subject identification, emotion perception, confidence rating, motion capture, facial emotion, facial expression, user study

ACM Reference Format:

Yun, C., Deng, Z., and Hiscock, M. 2009. Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. ACM Comput. Entertain. 7, 2, Article 21 (June 2009), 26 pages.
DOI = 10.1145/1541895.1541901 http://doi.acm.org/10.1145/1541895.1541901.

1. INTRODUCTION

In recent years, we witnessed an explosive growth in the creation and usage of high-fidelity 3D avatars in the entertainment industry, where the avatars are used as either representations of self or of others. The primary reason for the growth of high-fidelity avatars is due to growing audience expectations for more realistic avatars, as concomitantly the technology rises to meet the demand. As the technology pushes its limits further, it will become inevitable that the audience will have even higher expectations and demand more realistic avatars that truly resemble humans. The development of high-fidelity 3D avatars in entertainment and online environments is possible due to the use of motion-capture systems that allow developers to create 3D avatars and embed them with very detailed and realistic behaviors. Creating avatars whose body movements are very close to those of human beings has become more prevalent due to the relative ease with which the motions of the human body can be captured by recording actors' motions, processing their movements, and then implementing them in the avatars. For example, most current 3D sports game developers can capture the detailed athletic motions of actual athletes, so that the movements of the avatars in the games resemble very closely the actual body motions of the athletes. On the other hand, the usage of motion-capture systems to capture human facial motions (and/or emotions) is less prevalent due to the higher cost, effort, and time in processing and implementing them on the faces of avatars. In the movie industry, *Polar Express*. (http://polarexpressmovie.warnerbros.com/) and *Spiderman 3* (http://www.sonypictures.com/homevideo/spider-man3/) are two examples where a motion-capture system was used to capture facial motions and emotions. In the game industry, the demand for creating more precise high-fidelity 3D avatars that convey very detailed visual and behavioral realism of the face and body is even stronger than in the movie industry. For example, *Heavenly Sword* (http://www.us.playstation.com/heavenlysword/) and *Grand Theft Auto 4* (http://www.rockstargames.com/IV/) created avatars via motion-capture data that can move their bodies and faces in great detail. Unlike the movies, the main avatars (or main characters) in games represent the users interacting with other avatars (or characters), hence users demand that the avatars be capable of realistic actions and interactions closely resembling human-to-human interactions.

In terms of the high-fidelity 3D avatars created via facial motion-capture data, there is one issue that needs everyone's attention: that is, that the current audiences (or customers) of most entertainment media are no longer confined to one country. In other words, a medium (particularly a video game) that is

created in one country is generally not created for the audiences of that country exclusively. Rather, both producers and audiences anticipate that the medium will be available worldwide. Based on this, we must consider one seemingly obvious problem: if a medium is created in one country, can its high-fidelity avatars express emotions precisely and correctly to the audiences of other countries? Or more precisely, can the audiences from different nations and/or cultural groups perceive the emotions of the avatars as the avatars intended they should? Our study focuses on the ability of the high-fidelity 3D face avatar to deliver various types of emotions to different intercultural groups. For example, when motion-capture data of an actor from a particular nation or culture expresses various types of emotion, we are interested in observing whether the audience from the same nation or cultural background perceives these emotions more correctly than an audience from a different nation or cultural background, or will audiences from both groups see the emotions in the same way. The data would allow us to verify whether the national or cultural background of the expressors affect the perceptions of the perceivers. In conclusion, we will determine whether the locally generated high-fidelity 3D facial avatar is effective for global audiences, to help future entertainment industries create avatars that work well for global audiences.

Although there is no available research on high-fidelity 3D facial expression in an intercultural format, there are many studies available on the usage of avatars as embodied conversational agents (ECAs) in human computer interaction (HCI) applications [Andre et al. 1998; Cassell et al. 2000; Deng et al. 2006; Fabri et al. 2002; Gratch and Marsella 2006; Gratch et al. 2002; Lewis and Purcell 1984; Nass et al. 1998; Marsella and Gratch 2001; Rist et al. 1997; Ruttkay et al. 2002]. However, few studies have been conducted to measure the effectiveness of the avatars in intercultural settings where an avatar that was designed based on one cultural setting is presented to users from other cultural backgrounds. In particular, there is one area that lacks vigorous research on measuring the avatars' effectiveness: that is, *the emotional expressions generated by the avatars and the corresponding emotional perceptions and subject identifications by the users from different cultural backgrounds*. There is limited (if any) research that investigates the effectiveness of avatars with realistic 3D facial avatars embedded with high-fidelity emotion expressions.

In the field of psychology, the issue of whether facial expressions and the corresponding perceptions are universal or culturally specific has been debated for a long time, and considerable research has been done to prove (or disprove) the validity of one or the other. Some recent studies have shown the culturally specific nature of facial expressions and the way they are perceived by measuring the effectiveness of the perception of emotions by different cultural groups [Elfenbein and Ambady 2002, 2003; Elfenbein et al. 2002, 2007]. Their primary conclusion can be summarized in a simple phrase: "Familiarity breeds accuracy" [Elfenbein and Ambady 2003]. In other words, they concluded that when an expresser of emotion and the perceiver of an emotion have the same cultural background, the perceiver's recognition rate was higher than when they have different cultural backgrounds. However, other studies have shown universality in the expression of facial emotion and perception [Ekman 1994; Ekman and

Friesen 1971; Ekman et al. 1987; Matsumoto 2002, 2007] where the cultural background factor has little effect on the facial expression perception process. Although certain insights with regard to the relationship between emotion perception and cultural backgrounds are provided, these studies were mainly conducted with static photos that usually contain only a single highlighted frame of the entire expression process, rather than via animated avatars. Thus, it is still unclear whether the animated avatars will be able to produce similar results to the experiments conducted with static photos.

In the meantime, in the field of computer science, a few studies investigated via computer-generated avatars the relationship between the perception of emotion and cultural background. In their study, Bartneck et al. [2004] observed that the perception of expression and the culture of the perceivers of the expression were generally independent when two simple, iconic, avatar animations with various emotions were shown to participants. However, Bartneck et al. also identified some avatar animations that were perceived differently by participants from different cultures. On the other hand, the study by Koda and Ishida [2006] showed that the cultural backgrounds of the perceivers clearly affected how they perceived the emotions of comic/anime type avatars that expressed twelve different emotions (both human and nonhuman animations). These studies, similar to those by psychologists, still do not validate whether the cultural background affects the perception of the emotions when realistic 3D avatars embedded with high-fidelity facial expressions are used as a primary stimulus instead of the nonrealistic iconic or comic/anime drawn avatars.

In this study, we compare performance differences in subject identification, emotion perception, and confidence ratings among different cultural groups when they are shown clips of avatar animation in which very realistic 3D avatars are embedded with highly believable and culture-specific emotions.

The rapid advances in avatar-based technologies will inevitably force the entertainment industry to use even more realistic 3D avatars with highly believable emotions. Our attempt is to resolve whether realistic emotions expressed by avatars can be universally perceived by users from different cultural backgrounds. We intend to accomplish our goal by investigating whether culture-specific emotions have to be embedded on the avatar for different cultural groups or can the emotions from one cultural background be used effectively for other cultural groups.

In addition to the study on the perception of emotion, we also conducted a subject identification study to examine whether the perceivers were able to recognize the identity of the subjects who expressed the particular emotions embedded on the avatars. As in the emotion perception study, we aimed to find out whether the participants from the same background as the subjects achieved higher accuracy in identifying the subject than the groups from different cultural backgrounds did. We compared the accuracy of the subject identification with the confidence ratings to see whether there was any meaningful correlation between them.

In our experiment, we used a 3D facial model embedded with detailed, highly believable emotions to prove (or disprove) whether people of one cultural group could recognize the expressed emotions of their own cultural group more

Fig. 1. A VICON motion-capture system.

successfully than those of the other cultural groups. This was accompanied by a question that asked them how confident they were of their choices in identifying each emotion expressed by the avatar animations. To conduct this experiment, we created a realistic avatar using two principal components: the first comes from a scanned high-quality 3D facial model; the second component is a high-fidelity expressive facial motion data obtained by an optical motion-capture system (Figure 1). We captured the expressive facial motions of three chosen human subjects while they spoke three English sentences conveying five different emotions each. The first four emotions were anger, sadness, happiness, seriousness, and one of two emotions, either surprise or disgust. After acquiring the motion data, the captured expressive facial motion data was transferred to the pre-constructed static model of a 3D face to generate animated faces. In order to become familiar with how individuals uniquely express emotion and how different emotions are expressed in different ways, selected video clips of the subjects during the facial motion-capture process and the corresponding animated clips were shown simultaneously to the two groups of participants. Then the groups were shown only the animated faces, without voice, and asked to recognize the type of emotion expressed in each case. We also asked the participants to rate the confidence level of their decisions in each case to see whether the intensity of the facial expressions would impact the participants' confidence ratings as well as the accuracy of their perceptions of emotion.

In terms of experimental design, we broadly categorized the students as either US or international students, instead of dividing them into more nation/culture-specific groups. We are aware that there are hundreds of nationalities/cultures around the globe, and we consider our study as a first

step in this area. We believe our study will stimulate other researchers to continue this work.

The remainder of the article is organized as follows: Section 2 briefly reviews recent efforts that are most related to our work; Section 3 describes the set-ups and procedures of our experiment; Section 4 describes the results and corresponding conclusions; followed by discussion (Section 5); and future directions (Section 6).

2. RELATED WORK

A significant number of user studies have been conducted to evaluate human-like computer interfaces for a broad variety of applications [Bailenson and Yee 2006; Bente et al. 2001; Bonito et al. 1999; Busso et al. 2004; Guadagno et al. 2007; Hongpaisanwiwat and Lewis 2003; Katsyri et al. 2003; Koda and Maes 1996; Nowak and Rauh 2005; Sproull et al. 1996; Walker et al. 1994]. Among these studies, evaluating the avatars' usability and impact on task performance under various conditions has been a hot topic [Bente et al. 2001; Pandzic et al. 1999]. For example, researchers did user studies to evaluate the impact of using one's own face as the avatar [Nass et al. 1998] or to measure performance based on the level of visual realism of the avatars [Pandzic et al. 1999]. Currently, one of most debated issues is how the degree of realism conveyed by the avatars affects the usefulness of avatars in various applications, for which there are two possible outcomes. One is that the degree of realism conveyed by the avatars is independent of the avatars' usefulness. The study by Zanbaka et al. [2006] claims that there is a similarity in the avatars' persuasiveness, regardless of their degree of visual realism or even whether the avatars are human or nonhuman characters. The other outcome is that the degree of realism that avatars convey does affect the avatars' usefulness. A recent study based on a meta-analysis of 46 studies [Yee et al. 2007] concluded that the usefulness of more realistic human-like representation in social interaction settings was greater than less realistic counterparts. For researchers, findings like those of Yee et al. [2007] provide an essential reason for the importance of further work on the development of more realistic human avatars, their application, and subsequent evaluation. The conclusions of Yee et al. [2007], Garau et al. [2003], and Kang et al. [2008] also confirm that the combination of the avatars' high-quality visual realism and high-fidelity behavioral realism affects both perception and performance in a positive way. Our study focuses on the evaluation of highly realistic 3D facial avatars embedded with high-fidelity facial emotions for subject identification, emotion perception, and confidence ratings by perceivers from different cultural backgrounds.

There have been many studies that investigated the embedment of emotions on animated avatars. Some studies, such as the work of Fabri et al. [2002], used a generic face to express the emotion by exaggerating or manipulating the size of eyes, mouth, and eyelids. In contrast, there is the sophisticated design of a face made by applying expressive facial motion data to embed human emotions [Cassell et al. 2000]. Researchers also looked into effectiveness studies that used expressive motion-capture data to observe how people perceive emotions

[Cassell et al. 2000; Deng et al. 2006]. Two particular areas drew the attention of researchers: the degree of emotional intensity and that of intercultural background. Hess et al. [1997] investigated how the degree of intensity in emotional expression (20%, 40%, 60%, 80% and 100%) affects the accuracy of perception for four emotions (anger, disgust, sadness, and happiness) by using photographs as stimuli. The results show that the accuracy of perception for each emotion was affected linearly by the intensity of emotional expression. A similar study was conducted by Bartneck and Reichenbach [2005], while differentiating itself from the work of Hess et al. [1997] by using a cartoon-type human facial avatar as a stimulus. In this study, animated avatars with five different emotions (happiness, sadness, anger, fear and surprise) with ten different degrees of emotional intensity (10% to 100%) were shown to participants in order to rate the perceived intensity of the emotion as well as the difficulty in recognizing the emotion. A curve-linear relationship between the degree of intensity and the perception of intensity was observed. Meanwhile, the accuracy of recognition became negligible after the degree of intensity reached 30%. In the case of the study on difficulty, the participants rated the task as difficult when the degree of intensity was low and rated the task as easy when the degree of intensity was high. In addition, they rated fear as the most difficult emotion to identify. In our experiment, we created high-fidelity 3D avatars based on the emotions captured from three human subjects with three levels of intensity in facial expression (less expressive non-actor, expressive non-actor, and amateur actor). We then asked participants to identify the emotion and to rate how confident they were of their choices. We aimed to uncover how the intensity of facial expression affects accuracy in the perception of emotion.

How the expressive motions of nonrealistic avatars were perceived by participants from different cultural backgrounds was also studied [Bartneck et al. 2004; Koda and Ishida 2006]. In the study by Bartneck et al. [2004], Japanese and Dutch participants were asked to rate the degree of arousal and valence in the animations of two simple iconic avatars (created by a Japanese professor) with 30 different sets of emotions. Bartneck et al. found that, in general, the emotions were culturally independent in the area of perception, although some of the animations revealed that culture clearly affected perception. Koda and Ishida [2006] used 40 avatars (human figures, animals, plants, objects, and imaginary figures) created in a comic/anime drawing style by Japanese designers to observe whether participants from eight different countries (Japan, Korea, China, France, Germany, UK, USA, and Mexico) were able to identify 12 different facial expressions (happy, sad, approving, disapproving, proud, ashamed, grateful, angry, impressed, confused, remorseful and surprised). They found that the recognition accuracy for facial expressions by Japanese participants was the highest, and claimed that the cultural backgrounds of the participants affected the perception of the avatars' facial expressions. In addition, they found that the participants from Korea achieved the second-highest accuracy. Koda and Ishida [2006] interpreted this as the result of cultural similarity between Japan and Korea due to the spatial proximity of the two countries. This study concluded that cultural background does affect the perception of the avatars' facial expressions, as it does in the perception of human facial

expression. In both studies (Bartneck et al. [2004] and Koda and Ishida [2006]), iconic and nonrealistic human facial avatars were chosen as the primary stimuli to investigate whether the cultural background affects the identification of the avatars' emotions. However, neither study provides enough conclusive evidence that the perceived facial expressions of the avatars and the cultural backgrounds of the participants would remain independent if high-fidelity 3D human facial avatars were used as primary stimuli. In our study, the participants were divided into two groups: US and international. We wanted to see whether the perception of emotion is culturally independent in general, as claimed by Bartneck et al. [2004], or is affected by the cultural background of the participants, as in the study by Koda and Ishida [2006].

There are also several studies conducted by psychologists to verify whether accuracy in perceiving emotions is affected by cultural background. Some of the recent studies concluded that the cultural backgrounds of the expressers and the perceivers of emotions do affect accuracy in recognizing emotions [Elfenbein and Ambady 2002, 2003; Elfenbein et al. 2002; Yuki et al. 2007]. When the cultural background of the emotion expresser matches the cultural background of the perceiver, the resulting success rate is higher than when two are not from the same background. This is because individuals from the same cultural background are exposed to a greater degree to expressions of emotion that are generally accepted in that culture. Furthermore, greater exposure to the particular culture allowed the members of that culture to become familiarized with the socially approved emotion. The study by Yuki et al. [2007] went a step further by identifying how Americans and Japanese perceive emotions differently because they give greater or less emphasis to cues provided by facial expressions. The study confirmed that Japanese gave more weight to the eye area during the emotion perception process, while Americans gave more weight to the mouth area. Hence this study not only verified the cultural specificity in the perception of facial expression of emotions but also revealed the mechanism by which people in different cultures express and perceive facial emotions differently.

But there are also many studies that demonstrate that the emotion-perception process is universal, regardless of the cultural backgrounds of the expressers and perceivers [Beaupre and Hess 2005; Ekman 1994; Ekman and Friesen 1971; Ekman et al. 1987; Matsumoto, 2002, 2007]. Matsumoto [2002] pointed out the unwarranted nature of "in-group advantage in judging emotions across culture" from the meta-analysis of Elfenbein and Ambady [2002], based on the fact that the data was either under-qualified in terms of methodological requirements in examining in-group advantage or did not demonstrate in-group advantage after re-examination. In another study, he again, confirmed that the nationality of the expressers and perceivers did not affect the perception of emotion based on the observation that the perceivers did not alter their judgment of the emotion on the basis of the expressers' nationality [Matsumoto 2007]. Beaupre and Hess [2005] wanted to observe whether there were in-group advantages among French-speaking native Canadians, sub-Saharan Africans and Chinese immigrants in the expression of emotion and perception processes, and found that their data did not support an in-group advantage.

In all the studies mentioned so far, photographs were used as the stimuli for perceiving emotion. Katsuri [2006] concluded that the dynamic faces were perceived to express emotion better than the static photographs. In our study, high-fidelity 3D facial avatars were used as the stimuli to observe whether they had the same effect as static photos of real people in terms of the perception of emotion by different cultural groups.

The "forced-choice" experimental design was highlighted [Elfenbein and Ambady, 2003] for use in cross-cultural emotion communication studies where the perceivers respond from a predetermined list after viewing the expressions. Frank and Stennet [2001] offered flexibility in the "forced-choice" experimental design so that the perceivers had the option to choose "inconclusive" if they could not perceive the type of the emotion. Our study implements the unforced choice experimental design by providing the option of "unidentifiable" to the perceivers.

Several studies demonstrated that dynamics such as head motions and emotional transitions are also vital components in person identification. Two studies found that people were more accurate in identifying moving faces than still faces. Knight and Johnston [1997] showed participants two types of video footage: one was of still faces and the other of moving faces to test whether facial movements helped the participants to identify persons. They also showed the videos in normal (positive image) and degraded (negative image) forms, and found that participants were more successful in identifying persons when they were shown moving faces in degraded form. Lander et al. [2001] observed similar results when they degraded the video footage by pixelation or by blurring and showed the footage to the participants for person identification. Regardless of the degree of degradation by either pixelation or blurring, participants were more successful in identifying the moving face than the still face.

Katsyri and Sams [2008] performed a basic emotion perception study using synthetic and natural faces as stimuli, shown in either static or dynamic form. They observed that dynamic synthetic expressions were identified more successfully than static synthetic expressions, while natural expressions did not display any significant difference from dynamic and static faces. The stimuli of our study also include facial dynamics such as head motions and transitions in expression from neutral to specified emotions. We examine whether and how the dynamics affected our subject identification and emotion perception experiments in intercultural settings.

3. EXPERIMENTAL SET-UP

The purpose of this study is to investigate whether the national/cultural backgrounds of US and international participants affected their accuracy in subject identification and emotion perception in a setting where they were shown a series of high-fidelity 3D facial avatar animation clips. We also designed our study to observe the effects of the avatars' expressions and dynamic transitions on the groups. We first captured data from three human subjects who spoke in English and expressed designated emotions using a VICON motion-capture

system. After processed the data, we transferred it to a photorealistic 3D facial model and created a total of 45 facial animation clips. We chose 15 clips, arranged them randomly, and asked participants to complete a survey that includes questions concerning subject identity, emotion perception, and the level of confidence they had in their choices.

3.1 Facial Animation

To create high-fidelity 3D avatars with realistic facial expressions, we decided to capture the expressive facial motions of three human subjects who were chosen on the basis of the following criteria:

Subject 1 (S1): non-actor with low-intensity dynamics and expression;
Subject 2 (S2): non-actor with low-intensity dynamics and medium-intensity expression;
Subject 3 (S3): actor with high-intensity dynamics and expression.

The first two subjects, S1 and S2, were computer science graduate students, while the third subject (S3) was a senior majoring in theater and performing arts at the University of Houston. Their ages ranged from 25 to 30 ($M = 27.33$, $SD = 2.52$). We categorized S1 and S2 as low- and medium-expressive non-actors based on the degree of intensity of their facial expressions in a natural setting. We chose the non-actors because we wanted to capture facial expressions with a low- and medium-degrees of expression and dynamic intensity (which was their typical demeanor in normal circumstances). In relative terms, high-intensity expression was ranked as 100%, low-intensity expression as 25% to 35%, and medium-level intensity ranked from 60% to 70%. For dynamic intensity, we defined low intensity as low frequency and velocity head movements (or the near-absence of movements), while high intensity is defined as high frequency and velocity head movements. We excluded the sound quality (i.e., the volume of the subject's voice) from our criteria since our study is focused on nonverbal facial expressions.

We used a VICON motion-capture system to record high-fidelity expressive facial motions of the three subjects at a 120-Hz sampling frequency while an off-the-shelf digital video camera was used to simultaneously record the motion-capturing process as it focused on the subjects' faces from the front. A total of 10 VICON MX-40 cameras were used to capture facial expressions in detail. We placed 99 markers (each marker approximately 5mm in diameter) on feature points of each subject's face to capture their unique facial expressions as well as the subtle differences between the subjects (Figure 2). The facial marker layouts for the subjects were all the same. We also placed four head markers (each marker about 16mm diameter) on each subject's head to capture his/her head motions during the emotion-capture sessions. We asked each captured subject to speak three preselected English sentences five times each with corresponding emotions:

Sentence 1: "*This was easy for you*".
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), seriousness (SER), surprise (SUR)

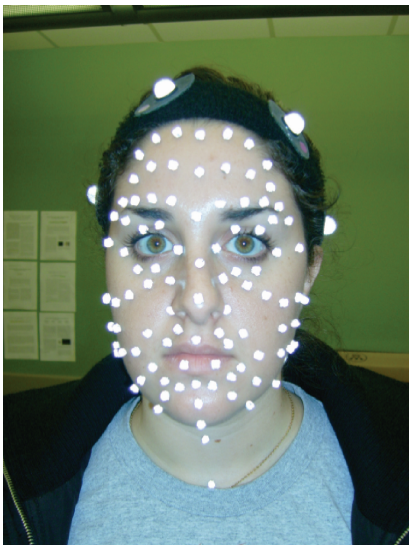


Fig. 2. A captured subject with facial markers.

Sentence 2: "*Those thieves stole thirty jewels*".
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), seriousness (SER), disgust (DIS)
Sentence 3: "*She is thinner than I am*".
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), seriousness (SER), surprise (SUR)

We intentionally chose short sentences to aid the subjects in expressing their emotions as naturally and realistically as possible. The subjects, especially S1 and S2 in particular, had difficulty in expressing and sustaining each emotion when the sentences were long.

In our study, we chose six emotions (anger, sadness, happiness, seriousness, surprise, and disgust). With the exception of the "serious" category, the emotions were based on the universal basic emotion classification [Ekman and Friesen 1971].

By using a high-fidelity optical capture system such as the VICON, it was possible to capture the distinct and detailed dynamics and characteristics of the facial expressions of different subjects. In addition, as an effort to create a realistic avatar, we considered beforehand how the avatar's eye motions had to be handled throughout the motion- capture process. Garau et al. [2001] demonstrated how the avatar's eye movements in a conversational instance impacted the quality of communication, so that the avatar with the gaze that conformed to the conversation outperformed the avatar with a random gaze. For our study, since the motion-capture process could not capture eye movement, we asked the subjects to gaze straight ahead while speaking during a motion-capture session. Fixing eye movement in a forward direction saved the avatar from possible adverse effects due to the negative impact of uncoordinated movements of the eyes and face.

We created beforehand a high-quality 3D facial model of a middle-aged American Caucasian female using a 3D scanner. We chose this model for our study for two reasons. First, we aimed to investigate how the subjects identified and perceived the emotions of a 3D avatar speaking American English with American-style facial expressions. So we wanted our visual stimulus be as close to an American as the language and facial expressions. The second reason is that in a previous study Zanbaka et al. [2006] showed a cross-gender influence whereby women were more influenced by male avatars and men were more influenced by female avatars. This cross-gender influence was observed regardless of whether the avatars were real humans, virtual humans or virtual characters. There were a total of 12 female and 46 male participants in our study. Hence, to enhance the performance of the participants, we decided to use the female avatar instead of the male.

We then transferred the captured facial motion data onto of the 3D facial model to synthesize the corresponding facial animations for each subject. This process allowed subjects' identities to be hidden behind the identical 3D model for later subject identification experiments. The process was similar to those of Knight and Johnston [1997] and Lander et al. [2001] where the true identities of the subjects were masked by processing their stimuli video clips. With the 3D model and the motion data, we created facial animation clips using Autodesk (formerly known as Alias) Maya software. Each clip contained a different subject speaking three sentences with different emotions. Similar to a study by Katsyri and Sams [2008], the animation clips retained the dynamic information of head motions and the transitions in facial expression (e.g., transition from initial neutral expression to designated expression in each piece of data).

We emphasized both the high-quality visual and the behavioral realism of the avatar in the creation of the animation clips. Although the avatar suffered some loss in the degree of visual realism during the process of creating the animation, the result still carried enough facial detail to be considered very realistic. Using the facial avatar animations and the footage from the motion-capture process of the subjects, we created two types of clips for the study. The first type contains the animation of the facial avatar and the corresponding recorded video sequence of each subject speaking during the motion-capture process (Figure 3). This was designed to get the participants familiarized with both subtle and distinct facial expressions for different emotional types expressed by different subjects. For example, the clip was able to show distinguishable characteristics when a single subject showed both the happy (HAP) and the angry expression (ANG), and when two different subjects had the same surprise (SUR) facial expressions. We showed movie clips of the motion-capture process instead of static photos because participants who became familiar with faces via movies rather than still pictures were more successful in identifying people [Roark et al. 2003]. The second type of clip contained an animation clip where the avatar spoke one of the sentences with one of the emotions listed above (Figure 4). Out of total 45 possible combinations, we chose 15 animation clips and arranged them randomly for the perception experiment.

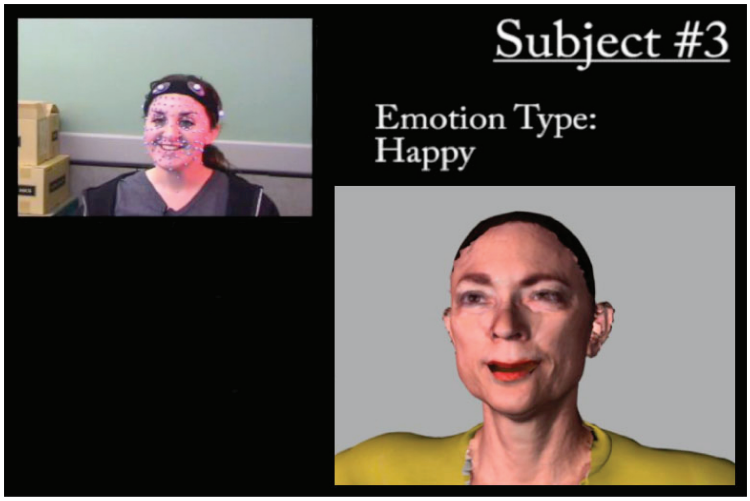


Fig. 3. Video of a subject during the motion-capture session, along with the animated face from motion-capture data to let survey participants become familiarized with the subject's expression of emotion.

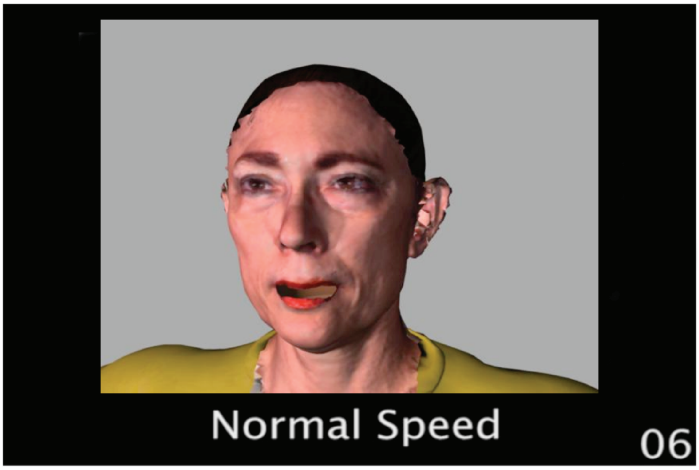


Fig. 4. The animated face alone, for subject identification and emotion perception.

3.2 Study

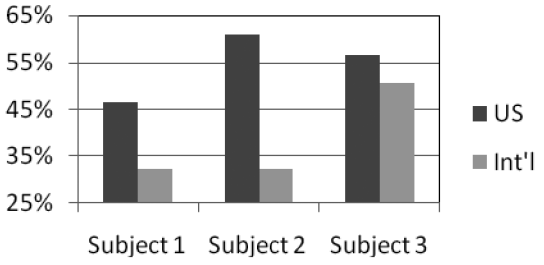
The participants of the survey were asked to perform the following three tasks: *Identify the subject, recognize the emotion type, and rate the confidence levels*. For the survey, we choose two groups of people based on following status:

Group 1 (G1): US students (24)
Group 2 (G2): International students (34)

Fifty-eight participants between the ages of 19 and 37 students at the University of Houston volunteered for the study. The students from various disciplines (i.e., computer science, engineering, and business majors) were divided into two groups based on their nationality (as in Bartneck et al.'s [2004] study): there were 24 US students and 34 international students. All the US students were born and lived in the United States, the group up was made up of 16 Caucasians, 5 Hispanics, 2 Asians, and 1 African-American, ranging in age from 19 to 34 ($M = 24.92$, $SD = 4.59$). Two out of 24 US students were female (8.38%). The international student group consisted of 17 Chinese, 5 Indians, 3 Mexicans, 2 Tunisians, and 1 person each from Jordan, Korea, Russia, Ukraine, Iran, Nigeria, and Indonesia, ranging from 22 to 37 years of age ($M = 25.63$, $SD = 4.36$). They lived in the US from 6 months to 2 years prior to this study. Ten out of 34 international students were female (29.41%). We formed these two groups (G1 and G2) to observe how they responded (either differently or similarly) to the US avatar animation clips in terms of subject identity, emotion perception, and confidence rating.

To prevent any potential interaction between the two groups, each group completed the survey at a separate location. They were also asked not to interact with one another within the group so as to obtain uninfluenced results.

We conducted our study using an unforced-choice method. The survey was conducted in a two-phase protocol. In the first phase, we handed out the survey and showed the first type of clip as a stimulus (containing the motion-capture processes of three subjects and the corresponding avatars) to the participants. The participants were asked to pay close attention to the unique subtleties and distinctions among the different subjects as they expressed various emotions. The stimuli were shown three consecutive times to familiarize the groups with the distinctions among the facial expressions and the subjects' unique ways of expressing emotions. We designed this familiarization process based on the study by Roark et al. [2003], where it was demonstrated that more face-learning experiences prior to the experiment on recognition improved the results of the experiment. In the second phase, the groups were shown the second type of clip containing 15 randomly selected animated avatar clips to answer a questionnaire composed of questions on subject identity, emotion type, and confidence rating. First, after viewing each clip, the participants were asked to identify the subject in the animated clip. They were asked to choose from one of the following four options: subject 1 (S1), subject 2 (S2), subject 3 (S3), or unidentifiable. Second, they were asked to perceive the emotion type in each animated clip and choose one from following seven options: Anger (ANG), sadness (SAD), happiness (HAP), seriousness (SER), disgust (DIS), surprise (SUR), or unidentifiable. And finally, they were asked to fill out confidence ratings. The rating was designed on a one-to-ten scale where one is "absolutely unconfident" and ten ranks as "absolutely/superbly confident". To aid the groups in making more detailed observations, each clip was played twice at a normal speed and then once at 33% of normal speed. Afterward, a 10-second pause was given for the participants to answer each survey question.



Subject	Subject 1	Subject 2	Subject 3
US	46.38%	60.87%	56.52%
Int'l	32.29%	32.29%	50.69%
Difference	14.09%	28.58%	5.83%

Fig. 5. Successful three-subject identification rate of US and International students.

4. RESULTS AND CONCLUSIONS

4.1 Subject Identity and Confidence Rating

In the subject identification study, we used a 2 (group) $\times$ 3 (subject) analysis of variance (ANOVA) to examine the accuracy with which groups identified the subject on whom each avatar's expressions was based. We found that the US student group (G1) performed better in the subject identification experiment than the international student group (G2), $F(1,159) = 9.00$, $p < .01$. When the avatar with the embedded expressions of non-actors (regardless of whether the non-actor was low-expressive (S1) or medium-expressive (S2)) was shown in the subject identification experiment, G1 had higher success in identifying the right subject (46.38% to 32.29% in identifying S1 and 60.87% to 32.29% in identifying S2). Figure 5 shows how each group performed overall in the subject identification experiment. The accuracy of G1 and G2 in identifying the actor (S3) was 56.52% and 50.69%, where the difference was comparatively low (5.83%). Based on these results, we can conclude that the rate of accuracy in subject identification is higher when the cultural background (or nationality) of the subjects and observers is identical. However, this is applicable only when the facial expressions data of real, everyday people is embedded on the avatar. When the facial expressions of actors are used, the advantage of cultural familiarity (or in-group advantage) becomes irrelevant. This implies that the entertainment media using the avatars with high behavioral realism will not cause the global audience to suffer the cultural unfamiliarity effect in identifying subject as long as the facial expressions of the avatars resemble those of the actors. In other words, the perceivers' cultural backgrounds will not affect the accuracy in identifying the subject if the behavioral intensity of the avatars' facial expressions is high enough to resemble the intensity of the facial expressions expressed by the actors. The difference in the number of participants between G1 and G2 was justified based on results that showed

that the performances of two groups were compatible in identifying the actor, while G1 performed with higher accuracy in identifying non-actors. There is an alternative way to explain why G1 and G2 performed at a similar level of accuracy in identifying S3, while G1 outperformed G2 in identifying S1 and S2. During the facial motion-capture processes, we observed almost no dynamics expressed by the non-actors, while the actor's dynamics were intense. The information related to the dynamics provided G2 sufficient information to identify S3 as well as G1 did. We conclude that the dynamics can also offset the cultural unfamiliarity disadvantage that G2 had on subject identification. On the other hand, since there were almost no dynamics in S1's and S2's motion data, the cultural unfamiliarity disadvantage influenced G2 more significantly in identifying S1 and S2. We verified this significant effect of subject (or more specifically the dynamic intensity of the subject) on subject identification, $F(2,159) = 7.79$, $p < .01$. And there was also a significant interaction between the group and the subject, $F(2,159) = 223.81$, $p < .01$. Therefore, we conclude that in the absence of dynamics information, the cultural familiarity advantage has a great affect on subject identification. However, the presence of dynamics can effectively offset the factors in the cultural unfamiliarity disadvantage. Hence as long as it is possible to embed substantial intensity in dynamics and facial expressions, very realistic 3D avatars in the entertainment media can be effectively deployed for audiences regardless of their cultural backgrounds. This finding implies that when highly believable 3D avatars are created, it is not necessary to embed each avatar with culturally specific facial expressions to accommodate global audiences with diverse cultural backgrounds.

Although the confidence rating was based on emotion perception, we compared it to the accuracy in subject identity to find any correlation between them. In general, the confidence ratings and accuracy in subject identification corresponded well to one another. In other words, higher confidence rating corresponded to higher success rates in identifying the subject. Figure 6 shows how each group rated in an overall confidence rating in subject identification. As expected, both G1 and G2 gave the highest confidence rating to S3, followed by S2 and S1, according to the degree of emotional facial expression (observed in a study by Bartneck and Reichenbach [2005]). We also concluded that the strong presence of dynamics in S3 contributed to G1 and G2 giving S3 the highest confidence rating. A 2 (group) $\times$ 3 (subject) ANOVA verifies this finding, $F(2,162) = 3.91$, $p < .05$. In the meantime, G1 rated the confidence ratings slightly higher than G2 in all cases. One might argue that G1 ranked the confidence ratings higher because they were familiar with the culturally accepted facial expressions in US culture [Elfenbein and Ambady 2003], and concluded that less familiarity with the culturally stereotypical emotional facial expressions led G2 to lower their confidence ratings as well as to perform less accurately in subject identification. On the other hand, G1, equipped with familiar culturally accepted stereotypical emotional facial expressions, ranked the confidence ratings higher, while performing better in subject identification than the international students. However, in our study, the difference in confidence rating between G1 and G2 was not significant enough ($F(1,162) = 0.25$, $p =$ ns) to support this argument. Finally, there was a highly significant

Subject	Subject 1	Subject 2	Subject 3
US	5.15	5.58	6.21
Int'l	4.98	5.33	6.08
Difference	0.17	0.26	0.13

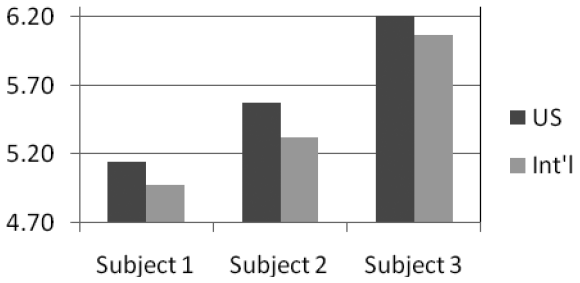


Fig. 6. Confidence rating for three subjects.

interaction between the group and subject ($F(2,162) = 260.21$, $p < 0.01$). Given the confidence rating and accuracy in subject identification, we can tell that G2 had a relatively difficult time identifying S2 and gave a lower confidence rating to S2 compared to G1 (the differences in ratings between G1 and G2 were 0.17 and 0.13 for S1 and S3, while it was 0.26 for S2).

4.2 Emotion Perception and Confidence Rating

For emotion perception, accuracy was assessed in a 2 (group) $\times$ 6 (emotion) ANOVA. We found that we had mixed results in the emotion perception experiment among US students (G1) and international students (G2), $F(1,336) = 0.88$, $p =$ ns. Figure 7 shows how each group performed overall in the emotion perception experiment. In perceiving sad (SAD), happy (HAP), serious (SER), and disgusted (DIS) emotions, G1 performed at least 7% more accurately than G2. Especially, in SER perception, G1 perceived more accurately than G2 by 18.75%. In perceiving anger (ANG) and surprise (SUR) emotions, G2 performed more accurately than G1 by 7.25% and 6.25%, respectively. Based on this data, we concluded that accuracy in emotion perception rates was not necessarily higher for every emotion type when the cultural background (or nationality) of the subjects and observers was identical. In terms of accurate perception of emotion among different emotion types, we observed significant differences, $F(5,336) = 19.74$, $p < .01$. The HAP emotion was the most successfully perceived emotion while the SAD emotion was the least successfully perceived emotion. In addition, there was also significant interaction between group and subject, $F(5,336) = 67.69$, $p < .01$. For example, other than the HAP emotion, G1 was most successful in perceiving DIS emotion while G2 was most successful in perceiving ANG emotion.

For a more in-depth analysis, we checked the emotion perception performance data from G1 and G2 where an avatar with a specific emotion type was shown with the expression data of three different subjects. We used HAP

Emotion	ANG	SAD	HAP	SER	DIS	SUR
US	43.33%	18.75%	66.67%	43.75%	54.17%	43.75%
Int'l	50.59%	8.82%	57.84%	25.00%	47.06%	50.00%
Difference	-7.25%	9.93%	8.82%	18.75%	7.11%	-6.25%

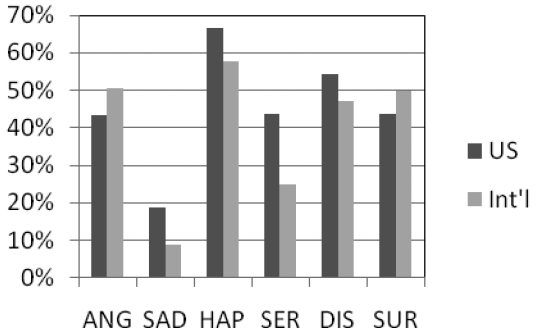


Fig. 7. Six successful emotion perception levels of US and international students.

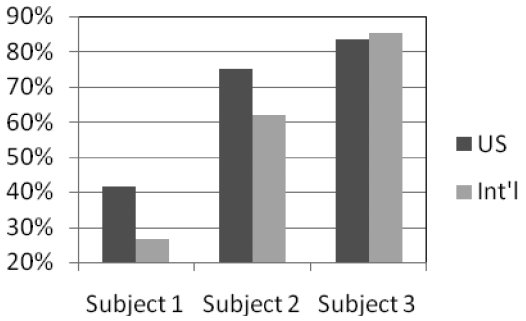


Fig. 8. Successful HAPPY emotion perception levels for US and international students.

emotion perception data to see how the groups performed when observing avatars with different intensities of emotion and dynamics. Figure 8 shows the difference in HAP emotion perception rates between G1 and G2. One obvious trend was the rise in the accuracy of emotion perception as the intensity of emotion and dynamics increased. As the groups saw avatar animations with higher intensities, they reached higher accuracy in the perception of the HAP emotion. Another subtle disposition was the impact of intensity in the accurate perception of emotion by G2. When the intensity of the emotion was in either a low or medium range, G2 performed worse than G1. However, when the intensity of the avatar's emotions was high, G1 and G2 performed almost identically. We conclude that G2 was more sensitive to the intensity of the avatar's expression than G1. In other words, people outside the avatar's country of origin will be more sensitive in perceiving an emotion type according to the intensity of expression and dynamics. In conjunction with the overall data, the data comparing the three different intensities of expression and the dynamics demonstrate

Emotion	ANG	SAD	HAP	SER	DIS	SUR
US	6.70	6.32	6.35	6.58	6.30	6.76
Int'l	6.05	5.74	5.63	5.32	5.81	6.30
Difference	0.65	0.58	0.73	1.25	0.48	0.45

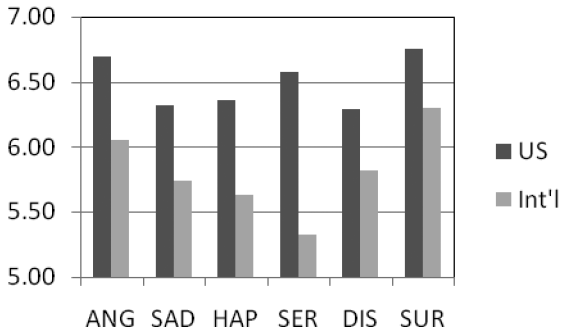


Fig. 9. Confidence ratings for six emotion types.

that the overall higher emotion perception by G1 was due to G2's lower success rate in perceiving emotions when the avatar's intensity of expression and dynamics was set to low or medium.

We surveyed the participants' confidence ratings to see how confident they were in their response to perceiving emotion. Here, a 2 (group) $\times$ 6 (emotion) ANOVA test demonstrated that neither the group ($F(1,324) = 0.51$, $p =$ ns) nor the emotion ($F(5,324) = 1.11$, $p =$ ns) was a significant factor in the confidence ratings. Figure 9 shows the participants' confidence ratings for six emotions. Group G1 gave higher confidence ratings for all six emotion types. This indicates that G1 at least had higher confidence in their decisions due to their familiarity with the types of expressions made by the avatars. In contrast, G2 gave a lower rating because the group was less familiar with the expressions. However, the difference in confidence ratings between the two groups was not significant. On the other hand, there was a significant interaction between group and emotion, ($F(5,324) = 138.43$, $p < .01$). Both groups gave ANG and SUR the two highest confidence ratings. Other than these two emotions, G1 and G2 had a different rating order: SER, HAP, SAD and DIS for G1 and DIS, SAD, HAP, and SER for G2. The degree of difference between the highest confidence rating and lowest confidence rating was more significant in G2 (6.76(SUR) – 6.30(DIS) = 0.99) than G1 (6.30(SUR) – 5.32(SER) = 0.46).

Comparing the data between the emotion perception and confidence ratings, we can observe some matching dispositions. For instance, by comparing the difference data from Figures 10 and 11, we can see that the data is similar for SAD, HAP, SER, and DIS emotions. The SER emotion confidence rating showed the biggest difference between G1 and G2, while DIS had the lowest difference rating for both emotion perception and confidence among the four emotions. However, we can see that G2 had higher accuracy in perceiving ANG and SUR emotions, even though G2 had lower confidence ratings than G1. Furthermore,

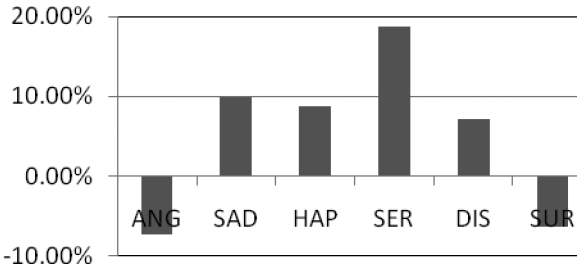


Fig. 10. Difference in accuracy in emotion perception between US and international students.

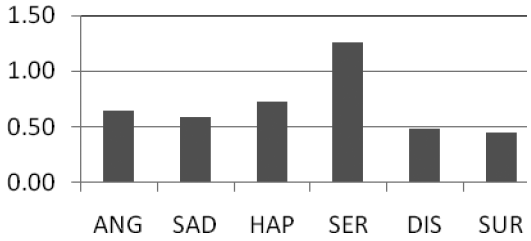


Fig. 11. Difference in confidence ratings between US and international students.

there are additional discrepancies between the emotion perception data and the confidence ratings. First, G1 and G2 rated their confidence in identifying the HAP emotion as one of lowest among six emotions. However, the actual emotion perception data demonstrates that both groups showed the highest accuracy in perceiving HAP emotion. Second, both groups gave almost similar confidence ratings for SAD and HAP emotions. However, the results show that both groups were least successful in perceiving the SAD emotion and most successful in perceiving the HAP emotion. Hence we must conclude that the discrepancy between emotion perception and confidence rating data is too obvious to support any agreement between them.

We also observed each group's confidence rating data when viewing the avatars embedded with the motion data of the different subjects. Here, we used HAP confidence rating data to observe how the groups performed when observing avatars with different intensities of emotion and dynamics. Figure 12 shows each group's confidence rating data. One obvious trend is the rise in confidence ratings as the emotion and dynamic intensity increases. As the groups observed higher-intensity-data-based avatar animations, they gave higher confidence ratings. Low intensity resulted in slightly lower confidence ratings. Another subtle factor was the impact that the level of intensity had on G2's confidence ratings. When the intensity of the emotion was in either a low or medium range, G2 rated lower than G1. However, when the intensity of the avatar's emotions was high, G2 gave a slightly higher confidence rating than G1. We conclude that G2 is more sensitive to the intensity of an avatar's emotion than G1. In other words, people outside the country from which the avatar's motion data was obtained will be more sensitive to the intensity of the emotion and dynamics. However, the confidence rating was not affected as severely

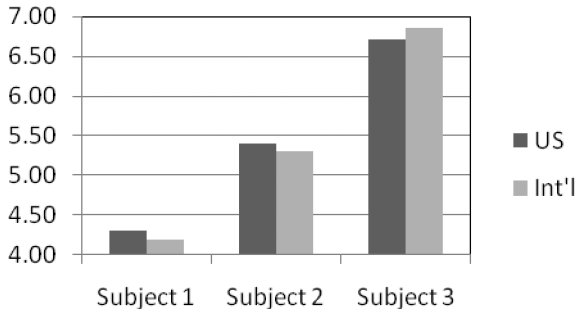


Fig. 12. Confidence ratings for the HAPPY emotion.

as emotion perception regarding sensitivity to intensity, since the differences between G1 and G2 in all three cases were no higher than 0.15.

5. DISCUSSION

Based on our results, we suggest a few guidelines for high-fidelity 3D avatar design and development to accommodate global audiences.

5.1 High Emotion/Dynamics Intensity

Most of current entertainment industry captures expression data from professional actors in creating high fidelity 3D avatars, guaranteeing that avatar animation retains high- intensity expression and dynamics. *Avatars with high-intensity expression and dynamics allow both the local and global audiences to achieve approximately equal levels in subject identification and emotion perception.* It is important to mention the importance of dynamics: High-intensity dynamics aids the global audience to recognize subjects and emotional types, since dynamics are also part of the process of emotion expression. Removing dynamics can lower the global audience's ability to recognize the subject and emotion types. Hence creating high-fidelity 3D avatar animation that embeds both high-intensity expression and dynamics is the right direction to take, since embedding only the expression data will reduce the ability of the global audience in identifying the subject and perceiving the emotion.

5.2 Low- and Medium-Intensity Emotion/Dynamics

We selected non-actors to create the avatar animations with low- to medium-intensity expressions and low-intensity dynamics data, since non-actors are more natural than actors in maintaining a low- or medium-level of intensity. In terms of subject identification, the international group (G2) performed identically, regardless of the different levels of intensity in expression (low in S1 vs. medium in S2) in identifying the non-actors while the dynamics intensity for both was equally low. This led us to infer that the difference in the intensity of expression alone does not contribute to changing the success rate of the global audience in subject identification, assuming the range of intensity remains at

a low to medium level. Therefore, increasing the dynamics intensity is necessary to assist the global audiences to correctly recognize the subject with a low- and/or medium-level expression intensity.

The results for emotion perception by G1 and G2 were mixed for the different emotion types. G1 was more accurate in perceiving the SAD, HAP, SER, and DIS emotions, while G2 was more accurate in perceiving ANG and SUR. Observing how each group performed in recognizing the HAP emotion on different subject types provided a hint as to why there were mixed results in emotion perception. We found that G1 and G2 had an almost identical level of success in emotion perception when viewing avatar animations embedded with the actor with high expression and dynamics intensity. On the other hand, G1 was more accurate in perceiving emotion than G2 when viewing avatar animations embedded with non-actors with low to medium expressive intensity and low dynamics intensity. Comparing the two non-actors demonstrates how low dynamics intensity suppresses the performance of G2 in the perception of emotion more severely than that of G1. Therefore, in designing avatar animations, it is necessary to include a higher level of dynamics intensity to help the global audience in subject identification and emotion perception.

5.3 Implementing Emotion into a Game

Due to time, money, and/or technical constraints, current non-player character (NPC) design does not provide NPCs with the ability to show emotion (via facial expression), but methods to do so have and are being explored. Zubek and Khoo [2002] added a capability to NPCs to enable them to chat with players and express different types of emotions by exchanging messages in a shooting game. In their synthetic character design framework, Kline and Blumberg [1999] designated emotion as one of four major components that influences believable synthetic characters in making realistic decisions and actions. Freeman [2004] introduced "emotioneering" techniques whereby a combination of scenario, storyline, and gameplay characteristics and character-to-character relationship-based techniques were used to add emotions to the game. In particular, techniques such as NPC interest, NPC deepening NPC character arc, and NPC rooting interest techniques were used to introduce and develop emotions for the NPCs in a core-level NPC design process. Most computer and video games follow the rules above in implementing the emotions in the NPCs. So emotions expressed by the NPCs are usually in terms of immediate actions (e.g., attack the player on sight), relationships (e.g., become close friends), and attitudes (e.g., fear) toward the player character (PC) rather than in terms of facial expression.

In addition, the author also suggests using cinematic movie clips as a technique to add (or reveal) the emotions for both PCs and NPCs. Games such as *Heavenly Sword* and *Grand Theft Auto 4* are good examples that apply this technique with detailed facial expressions for both PCs and NPCs. However, few games are capable of letting the NPCs make the facial expressions. Even games that implement facial expressions use only their generic forms. For example, the *Elder Scroll IV: Oblivion*, one of Bethesda Softworks'

RPG games for Microsoft Xbox 360 video game console and 2006 Game of the Year (http://www.bethsoft.com/eng/games/games_oblivion.html), implemented a dialog system with NPCs where, depending on the topic the player chooses to speak about, the NPCs love, like, dislike, or hate the player. These four emotions are revealed via their respective facial expressions, which are, however, in low-detail. Furthermore, the differences between the two positive expressions (love and like) and the two negative ones (hate and dislike) were designed by controlling the emotional intensity. If we regard the love expression as a generic happy expression with 100% intensity, we can regard the like expression as a generic happy expression with about or less than 50% expression. Likewise, if we regard the hate expression as a generic anger expression with 100% intensity, we can regard the dislike expression as a generic anger expression about or less than 50% expression (it sometimes resembles a medium-intensity sad expression). As future work we are interested in designing the NPCs with detailed facial expressions in a computer game and investigating whether the results from the present study is valid during gameplay as well.

6. FUTURE WORK

Contrary to previous work by Elfenbein et al. [2002] and Elfenbein and Ambady [2003], our results show that the group of international students performed as accurately as the US students in the category of emotion perception. This result agrees with the work of Bartneck et al. [2004] which states that the recognition of emotion is independent of culture. To further confirm our findings, we plan to conduct more comprehensive user studies under a similar setting with an increased but equal number of US and international students. We also plan to have an equal number of male and female participants in order to observe how the two genders perform. Under the revised conditions, we will observe whether the unexpected results that we obtained in this study can be reproduced (i.e., the participants rated the confidence rating for the HAP emotion relatively low, although they were most successful in perceiving this emotion). In addition, we plan to implement our findings into a video game environment where the characters (both players and nonplayers) can form facial expressions during the game. We are also interested in splitting the group of international students into several subgroups based on their specific cultural backgrounds (Indians, Asians, English-speaking Europeans, and non-English-speaking Europeans) who lived in the US for a specific amount of time (6 months to a year) to see whether we could reproduce the results of Elfenbein and Ambady [2003]. This will show us how national- and culture-specific audiences perform differently. Finally, we are interested in doing facial motion-capture with a diverse cultural group other than from the US (e.g., Chinese). We want to use this dataset to conduct a similar experiment with several national/cultural subgroups (Americans, Chinese, nonChinese Asians, and Europeans).

REFERENCES

ANDRE, E., RIST, T., AND MULLER, J. 1998. Guiding the user through dynamically generated hypermedia presentations with a life-like character. *IUI '98*, 21–28.

BAILENSON, J. N. AND YEE, N. 2006. A longitudinal study of task performance, head movements: Subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments 15*, 6, 699–716.

BARTNECK, C., TAKAHASHI, T., AND KATAGIRI, Y. 2004. Cross-cultural study of expressive avatars. In *Proceedings of the Social Intelligence Design*, 21–27.

BARTNECK, C. AND REICHENBACH, J. 2005. Subtle emotional expressions of synthetic characters. *Int. J. Human-Computer Studies 62*, 2, 179–192.

BEAUPRE, M. G. AND HESS, U. 2005. Cross-cultural emotion perception among Canadian ethnic groups. *J. Cross-Cultural Psychology 36*, 3, 355–370.

BENTE, G., KRAMER, N. C., PETERSON, A., AND DE RUITER, J. P. 2001. Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *J. Nonverbal Behavior 25*, 3, 151–166.

BONITO, J. A., BURGOON, J. K., AND BENGTSSON, B. 1999. The role of expectations in human-computer interaction. In *Proceedings of GROUP '99: International Conference on Supporting Group Work*, 229–238.

BUSSO, C., DENG, Z., YILDIRIM, S., BULUT, M., LEE, C. M., KAZEMZADEH, A., LEE, S., NEUMANN, U., AND NARAYANAN, S. 2004. Analysis of emotion perception using facial expressions, speech and multimodal information. In *Proceedings of ACM 6th International Conference on Multimodal Interfaces* (ICMI 2004). ACM, New York, 205–211.

CASSELL, J., SULLIVAN, J., PREVOST, S., AND CHURCHILL, E. F. 2000. *Embodied Conversational Agents*. MIT Press, Cambridge, MA.

DENG, Z., BAILENSON, J., LEWIS, J. P., AND NEUMANN, U. 2006. Perceiving visual emotions with speech. In *Proceedings of the 6th International Conference on Intelligence Virtual Agents* (IVA) 4133, 107–120.

ELFENBEIN, H. A., LEVESQUE, M., BEAUPRE, M., AND HESS, U. 2007. Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion 7*, 1, 131–146.

ELFENBEIN, H. A. AND AMBADY, N. 2003. When familiarity breeds accuracy: Cultural exposure and facial emotion perception. *J. Personality Social Psychology 85*, 2, 276–290.

ELFENBEIN, H. A., MANDAL, M. K., AMBADY, N., HARIZUKA, S., AND KUMAR, S. 2002. Cross-cultural patterns in emotion perception: Highlighting design and analytical techniques. *Emotion 2*, 1, 75–84.

ELFENBEIN, H. A. AND AMBADY, N. 2002. Is there an in-group advantage in emotion perception? *Psychological Bull. 128*, 2, 243–249.

EKMAN, P. 1994. Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bull. 115*, 2, 268–287.

EKMAN, P. AND FRIESEN, W. V. 1971. Constants across cultures in the face and emotion. *J. Personality Social Psychology 17*, 2, 124–129.

EKMAN, P., FRIESEN, W. V., O'SULLIVAN, M., DIACOYANNI-TARLATZIS, I., KRAUSE, R., PITCAIM, T., SCHERER, K., CHAN, A., HEIDER, K., LCCOMPTE, W. A., RICCI-BITTI, P. E., AND TOMITA, M. 1987. Universal and cultural differences in the judgments of facial expressions of emotion. *J. Personality Social Psychology 53*, 4, 712–717.

FABRI, M., MOORE, D. J., AND HOBBS, D. J. 2002. Expressive agents: Non-verbal communication in collaborative virtual environments. In *Proceedings of Autonomous Agents and Multi-Agent Systems (Embodied Conversational Agents).*

FRANK, M. G. AND STENNETT, J. 2001. The forced-choice paradigm and the perception of facial expressions of emotion. *J. Personality Social Psychology 80*, 1, 75–85.

FREEMAN, D. 2004. *Creating Emotion in Games: The Craft and Art of Emotioneering*. 1st ed., New Riders Publishing.

GARAU, M., SLATER, M., BEE, S., AND SASSE, M. 2001. The Impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 309–316.

GARAU, M., SLATER, M., VINAYAGAMOORTHY, V., BROGNY, A., STEED, A., AND SASSE, M. A. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the CHI'03 Conference*.

GRATCH, J. AND MARSELLA, S. 2005. Evaluating a computational model of emotion. *J. Autonomous Agents Multi-Agent Syst. 11*, 1, 23–43.

GRATCH, J., RICKEL, J., ANDRE, E., CASSELL, J., PETAJAN, E., AND BADLER, N. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Syst. 17*, 4, 54–63.

GUADAGNO, R. E., BLASCOVICH, J., BAILENSON, J. N., AND MCCALL, C. 2007. Virtual humans and persuasion: The Effects of Agency and Behavioral Realism. *Media Psychology*, *10*, 1, 22.

KANG, S., WATT, J., AND ALA, S. 2008. Communicators' perceptions of social presence as a function of avatar realism in small display mobile communication device. In *Proceedings of the Hawaii International Conference on System Science*.

HESS, U., BLAIRY, S., AND KLECK, R. E. 1997. The intensity of emotional facial expressions and decoding accuracy. *J. Nonverbal Behavior 21*, 4, 241–257.

HONGPAISANWIWAT, C. AND LEWIS, M. 2003. Attentional effect of animated character. In *Proceedings of the Human-Computer Interaction* (IFIP INTERACT03), 423–430.

KATSYRI, J. 2006. Human recognition of basic emotions from posed and animated dynamic facial expressions. Ph.D. dissertation, Helsinki University of Technology.

KATSYRI, J., KLUCHAREV, V., FRYDRYCH, M., AND SAMS, M. 2003. Identification of synthetic and natural emotional facial expression. In *Proceedings of the International Conference on Auditory-Visual Speech Processing* (AVSP'2003), 239–244.

KATSYRI, J. AND SAMS, N. 2008. The effect of dynamics on identifying basic emotions from synthetic and natural faces. *Int. J. Human-Computer Studies 66*, 4, 233–242.

KLINE, C. AND BLUMBERG, B. 1999. The art and science of synthetic character design. In *Proceedings of the AISB 1999 Symposium on AI and Creativity in Entertainment and Visual Art*.

KNIGHT, B. AND JOHNSON, A. 1997. The role of movement in face recognition. *Visual Cognition 4*, 3, 265–273.

KODA, T. AND MAES, P. 1996. Agents with faces: The effect of personification. In *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication* (RO-MAN'96), 189–194.

KODA, T. AND ISHIDA, T. 2006. Cross-cultural study of avatar expression interpretations. In *Proceedings of the 2006 International Symposium on Applications and Internet* (SAINT 2006), 130–136.

LANDER, K., BRUCE, V., AND HILL, H. 2001. Evaluating the effectiveness of pixelation and blurring on masking theidentity of familiar faces. *Applied Cognitive Psychology 15*, 1, 101–116.

LEWIS, J. AND PURCELL, P. 1984. Soft machine: A personable interface. In *Proceedings of the Graphics Interface*, 223–226.

MARSELLA, S. AND GRATCH, J. 2001. Modeling the interplay of plans and emotions in multi-agent simulations. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*.

MATSUMOTO, D. 2007. Emotion judgments do not differ as a function of perceived nationality. *Int. J. Psychology 42*, 3, 207–214.

MATSUMOTO, D. 2002. Methodological requirements to test a possible in-group advantage in judging emotions across the cultures: Comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bull. 128*, 2, 236–242.

NASS, C., KIM, E. Y., AND LEE, E. J. 1988. When my face is the interface: An experimental comparison of interacting with one's own face or someone else's face. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 148–154.

NOWAK, K. L. AND RAUH, C. 2005. The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *J. Computer-Mediated Communication 11*, 1, 153–178.

PANDZIC, I. S., OSTERMANN, J., AND MILLEN, D. 1999. User evaluation: Synthetic talking faces for interactive services. *The Visual Computer 15*, 330–340.

RIST, T., ANDRE, E., AND MULLER, J. 1997. Adding animated presentation agents to the interface. In *Proceedings of the 2nd International Conference on Intelligent User Interfaces* (IUI '97), 79–86.

ROARK, D. A., O'TOOLE, A. J., AND ABDI, H. 2003. Human recognition of familiar and unfamiliar people in naturalistic video. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures* (AMFG '03), 36–41.

RUTTKAY, Z., DORMANN, C., AND NOOT, H. 2002. Evaluating ECAs – What and how? In *Proceedings of the AAMAS02 Workshop on Embodied Conversational Agents*.

SPROULL, L., SUBRAMANI, M., KIESLER, S., WALKER, J. H., AND WATERS, K. 1996. When the interface is a face. *Human-Computer Interaction 11*, 2, 97–124.

WALKER, J. H., SPROULL, L., AND SUBRAMANI, R. 1994. Using a human face in an interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing System: Celebrating Interdependence*, 85–91.

YEE, N., BAILENSON, J. N., AND RICKERTSEN, K. 2007. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1–10.

YUKI, M., MADDUX, W. W., AND MASUDA, M. 2007. Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J. Experimental Social Psychology 43*, 303–311.

ZANBAKA, C., GOOLKASIAN, P., AND HODGES, L. 2006. Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1153–1162.

ZUBEK, R. AND KHOO, A. 2002. Making the human care: On building engaging bots. In *Proceedings of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment.*

Received November 2008; revised January 2009; accepted February 2009